Notes and comments

Variation in genetic identity within kinships

WILLIAM G. HILL

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, U.K.

In a recent paper of this journal, Rasmuson (1993) computed the standard deviation (s.d.) of the genetic identity of relatives, i.e. of the proportion of the genome which is identical by descent. For example, for parent and offspring the mean identity is 50 per cent and its s.d. is 0 per cent because the offspring inherits exactly one-half of the autosomal genes from each parent; whereas for grandparent and grandoffspring or for half-sibs, the mean is 25 per cent and, as there is segregation and recombination, the s.d. is not zero, Rasmuson giving a value of 2.5 per cent for each case.

The variance of identity was computed by Rasmuson on the assumption that the genotype comprised a number of independently segregating discrete units, the segregation index, which was set at 100, a value considered appropriate for human females (equal to the haploid chromosome number plus the total number of chiasmata). In her model it is, in effect, assumed that the genome comprises 100 chromosomes in each of which no recombination occurs, whereas recombination events are essentially continuously distributed along the chromosomes and occur at different locations in each generation.

An equivalent problem to that considered by Rasmuson for the case of unilateral descendants is the variance of the contribution of the non-recurrent parental line in a backcrossing programme. Coincidentally, I have recently derived a formula to compute this variance for a model that takes account of the distribution of recombinations (Hill, 1993), and which is based on results of Franklin (1977).

Let us assume there are v chromosomes, with the jth of map length l_j and the total length L Morgans, and that recombination events are continuously distributed on each chromosome with no interference, in accordance with Haldane's (1919) mapping function. The proportion of the genome which is identical between relatives that are mth degree direct descendants (e.g. $m = 2$ for grandoffspring), P_m, then has mean $E(P_m) = 1/2^m$ and variance

$$V(P_m) = (1/2L^2)(1/4^m) \sum_{i=1}^{m-1} \binom{m-1}{i}(1/i^2)$$
$$\times [2iL - v + \sum_{j=1}^{v} \exp(-2il_j)]$$

for $m > 1$, and $V(P_1) = 0$ (from Hill, 1993).

The examples I computed for illustration (in livestock) were for $v = 30$ chromosomes of total map length 30 Morgans, 10 each of lengths 0.5, 1.0 and 1.5. This gives s.d. $(P_2) = 3.51$ per cent, in contrast to Rasmuson's value of 2.5 per cent (also obtained by taking $v = 100$ and $l_j \to 0$ in the equation). For fourth-degree descendants, $E(P_4) = 6.25$ per cent, and the corresponding s.d.s are 2.04 per cent and 1.65 per cent. A greater difference would apply with shorter chromosomes.

The variances are not the same for collateral relatives of corresponding mean identity. The reason is that, for two loci with recombination fraction c, the probability that a grandoffspring inherits genes at both loci from a grandparent is $(1 - c)/2$, whereas the probability that two half-sibs inherit the same pair of genes is $(1 - c)^2/2 + c^2/2$. Only for $c = 0$ (complete linkage) and $c = 1/2$ (no linkage) are these quantities equal. For half-sibs, the mean identity is $E(P_{HS}) = 1/4$ and the variance is

$$V(P_{HS}) = (1/128L^2)[4L - v + \sum_{j=1}^{v} \exp(-4l_j)]$$

using results of Franklin (1977). In our example s.d. $(P_{HS}) = 2.82$ per cent, lower than s.d. (P_2) for grandoffspring–grandparent. For full-sibs the variance is $2V(P_{HS})$, because the two parents' contributions are independent. Formulae can be computed for other relationships, but are more complicated and beyond the scope of this note. The methods of Weir and Cockerham (1969) could be employed, and also means and variances of identity of genotype derived for full-sibs, for example, rather than of genes as here.

Although these results which take account of the continuous structure of chromosomes do not differ qualitatively from those of Rasmuson, it seems preferable to use the more complete analysis.

References

FRANKLIN, I. R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theor. Pop. Biol.*, **11**, 60–80.

HALDANE, J. B. S. 1919. The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.*, **8**, 299–309.

HILL, W. G. 1993. Variation in genetic composition in backcrossing programs. *J. Hered.*, **84**, 212–213.

RASMUSON, M. 1993. Variation in genetic identity within kinships. *Heredity*, **70**, 266–268.

WEIR, B. S. AND COCKERHAM, C. C. 1969. Pedigree mating with two linked loci. *Genetics*, **61**, 923–940.